

OPTIMAL ON-LINE SELECTION OF AN ALTERNATING SUBSEQUENCE: A CENTRAL LIMIT THEOREM

ALESSANDRO ARLOTTO AND J. MICHAEL STEELE

ABSTRACT. We analyze the optimal policy for the sequential selection of an alternating subsequence from a sequence of n independent observations from a continuous distribution F , and we prove a central limit theorem for the number of selections made by that policy. The proof exploits the backward recursion of dynamic programming and assembles a detailed understanding of the associated value functions and selection rules. The methods used here suggest a profitable approach to the asymptotic analysis of other Markov decision problems.

KEY WORDS: Bellman equation, on-line selection, Markov decision problem, dynamic programming, alternating subsequence, central limit theorem, non-homogeneous Markov chains.

AMS SUBJECT CLASSIFICATION (2010): Primary: 60C05, 60G40, 90C40; Secondary: 60F05, 90C27, 90C39

1. INTRODUCTION

In a typical on-line selection problem, a decision maker is presented with n random values in sequence and must decide whether to accept or reject each newly presented value. In the most famous such problem, the decision maker gets to make only a single choice, and the goal is to maximize the probability that the selected value is the best out of all n values. A problem of this kind was considered by Cayley (1875), but the modern developments begin in the 1960's with notable work by Lindley (1961) and Dynkin (1963). Samuels (1991) gives a thoughtful survey of much of the earlier literature, and connections to more recent work are given by Krieger and Samuel-Cahn (2009), Buchbinder, Jain and Singh (2010), and Bateni, Hajiaghayi and Zadimoghaddam (2010).

In some more combinatorial problems, the decision maker makes multiple sequential selections from the sequence of presented values, and the objective is to maximize the expected number of selected elements, subject to appropriate combinatorial constraints. For example, one can consider the optimal sequential selection of a monotone increasing subsequence. This on-line selection problem was first studied in Samuels and Steele (1981), and it has been analyzed more recently in Gnedin (1999; 2000*a*; 2000*b*), Baryshnikov and Gnedin (2000), and Bruss and Delbaen (2001). The present investigation is particularly motivated by the work of

A. Arlotto: The Fuqua School of Business, Duke University, Durham, NC, 27708. Email address: alessandro.arlotto@duke.edu.

J. M. Steele: The Wharton School, Department of Statistics, Huntsman Hall 447, University of Pennsylvania, Philadelphia, PA 19104. Email address: steele@wharton.upenn.edu.

Bruss and Delbaen (2004) which establishes a central limit theorem for the sequential selection of a monotone subsequence when the number N of values offered to the decision maker is a Poisson random variable that is independent of the sequence X_1, X_2, \dots of independent random variables with common continuous distribution.

Here, we consider the problem of making on-line selection of an *alternating subsequence*:

$$X_{i_1} > X_{i_2} < X_{i_3} > \dots \geq X_{i_k} \quad \text{with } 1 \leq i_1 < i_2 < \dots < i_k \leq n.$$

To be completely explicit, we consider the class Π of Markov deterministic policies that are adapted to the sequence $\{X_i : 1 \leq i \leq n\}$, so the decision to accept or reject the value X_i that is offered at time i is a deterministic function of the vector (X_1, X_2, \dots, X_i) . One might also consider randomized policies, but, standard results in dynamic programming confirm that there is no profit in doing so here (cf. Bertsekas and Shreve, 1978, Corollary 8.5.1).

Given a policy $\pi \in \Pi$, we denote by $A_n^o(\pi)$ the number of selections made by π for the realization $\{X_1, X_2, \dots, X_n\}$. It was found in Arlotto, Chen, Shepp and Steele (2011) that for each n there is a unique policy $\pi_n^* \in \Pi$ such that

$$\mathbb{E}[A_n^o(\pi_n^*)] = \sup_{\pi \in \Pi} \mathbb{E}[A_n^o(\pi)],$$

and it was also found there that the behavior of the mean is very well controlled; specifically one has

$$(1) \quad \mathbb{E}[A_n^o(\pi_n^*)] = (2 - \sqrt{2})n + O(1).$$

Here our main goal is to show that $A_n^o(\pi_n^*)$ satisfies a central limit theorem.

Theorem 1 (Central Limit Theorem for Optimal On-line Alternating Selection). *There is a constant $0 < \sigma^2 < \infty$ such that*

$$\frac{A_n^o(\pi_n^*) - (2 - \sqrt{2})n}{\sqrt{n}} \implies N(0, \sigma^2) \quad \text{as } n \rightarrow \infty.$$

The value of σ^2 is not known exactly, but we give a formula that expresses σ^2 as infinite series. Monte Carlo calculations¹ suggest that $\sigma^2 \sim 0.3096$, but the determination of an explicit (closed-form) expression for σ^2 remains an open problem. It may even be a tractable one, though it is unlikely to be easy.

NATURE AND ORGANIZATION OF THE ANALYSIS

Our proof of Theorem 1 rests on the sustained investigation of the value functions that are determined by the recursive Bellman equation for our problem. The first implication of this analysis is that the optimal policy π_n^* is a threshold policy, i.e. the policy is characterized completely by a set $\{g_n, g_{n-1}, \dots, g_1\}$ of time-dependent threshold functions that tell us when to accept or reject a newly presented value. An early step in our investigation was the numerical calculation of these threshold functions, and we would have had little chance to develop the argument given here without the detailed guidance that is summarized in Figure 1.

Section 2 recalls those results from earlier work that are needed to make the present arguments self-contained. The main fact we need is the convenient form

¹Numerical estimates are obtained discretizing the state space with a grid size of 10^{-4} and performing $5 \cdot 10^5$ repetitions. The standard error for the variance equals 6.19×10^{-4} .

(3) of the Bellman equation, but we also make essential use of a technical property of the value functions that is given by (4), which we call *restricted supermodularity*.

Sections 3 through 6 develop the geometry of the threshold functions. Here we have been as systematic as possible so that one might see the features of our analysis that might carry over to other Markov decision problems. Roughly speaking, one frames concrete hypotheses based on the suggestions of Figure 1, and one proves these hypotheses by inductions that are assisted by two flavors of supermodularity. The specific inferences are particular to the problem of alternating selections, but the general pattern should apply to many Markov decision problems.

Sections 7 and 8 exploit the geometrical characterization of the threshold functions to obtain information about the distribution of $A_n^o(\pi_n^*)$, the number of selections made by the optimal policy for the problem with time horizon n . The main step here is the introduction of a horizon-independent policy π_∞ that is determined by the limit of the threshold functions that define π_n^* . It is relatively easy to check that the number of selections $A_n^o(\pi_\infty)$ made by this policy is a Markov additive functional of a stationary, uniformly ergodic, Markov chain. One can use off-the-shelf results to confirm that the central limit theorem holds for $A_n^o(\pi_\infty)$, provided that one shows that the variance of $A_n^o(\pi_\infty)$ is not $o(n)$. We then complete the proof of Theorem 1 by showing that there is a coupling under which $A_n^o(\pi^*)$ and $A_n^o(\pi_\infty)$ are close in L^2 ; specifically we show $\|A_n^o(\pi^*) - A_n^o(\pi_\infty) - \mathbb{E}[A_n^o(\pi^*) - A_n^o(\pi_\infty)]\|_2 = o(\sqrt{n})$.

2. THE BELLMAN EQUATION AND THE OPTIMAL POLICY

Since the distribution F is continuous and since the problem is unchanged if we replace X_i by $U_i = F^{-1}(X_i)$, we can assume without loss of generality that the X_i 's are uniformly distributed on $[0, 1]$. To make our analysis self-contained (and to avoid repetition), we need to recall a few facts from Arlotto, Chen, Shepp and Steele (2011).

First, there is a sequence $\{g_n, g_{n-1}, \dots, g_1\}$ of functions $g_k : [0, 1] \rightarrow [0, 1]$ so that, if we set $Y_0 \equiv 0$ and define Y_i recursively for $1 \leq i \leq n$ by

$$Y_i = \begin{cases} Y_{i-1} & \text{if } X_i < g_{n-i+1}(Y_{i-1}) \\ 1 - X_i & \text{if } X_i \geq g_{n-i+1}(Y_{i-1}), \end{cases}$$

then the number of selections made by the optimal on-line policy π_n^* is given by

$$(2) \quad A_n^o(\pi_n^*) = \sum_{i=1}^n \mathbb{1}(X_i \geq g_{n-i+1}(Y_{i-1})).$$

We also need a few facts about the value functions $v_k : [0, 1] \rightarrow \mathbb{R}^+$ that are defined for $1 \leq k \leq n$ by the expected sum

$$v_k(y) = \mathbb{E} \left[\sum_{i=n-k+1}^n \mathbb{1}(X_i \geq g_{n-i+1}(Y_{i-1})) \mid Y_{n-k} = y \right].$$

In words, $v_k(y)$ is the expected number of selections *yet to be made* by the optimal policy π_n^* when the number of observations yet to be seen is k , and the current state Y_{n-k} equals y . In particular, we have $v_n(0) = \mathbb{E}[A_n^o(\pi_n^*)]$.

The optimality principle of dynamic programming tells us that the value functions $v_k(\cdot)$, $1 \leq k < \infty$, can be recursively determined. Specifically, if we set

$v_0(y) \equiv 0$ for all $y \in [0, 1]$, then the functions $v_k(\cdot)$, $1 \leq k < \infty$, satisfy the Bellman equation

$$(3) \quad v_k(y) = y v_{k-1}(y) + \int_y^1 \max \{v_{k-1}(y), 1 + v_{k-1}(1-x)\} dx.$$

One implication of this recursion is that the value functions $v_k(\cdot)$, $1 \leq k < \infty$, are continuous and differentiable.

The last fact we need from our earlier analysis is that the value functions have a certain *restricted supermodularity* property. Specifically, they satisfy the inequality

$$(4) \quad v_k(u) - v_k(1-y) \leq v_{k+1}(u) - v_{k+1}(1-y) \text{ for all } y \in [0, 1/2] \text{ and } u \in [y, 1-y].$$

This property may seem technical, but one cannot do without it. It provides an essential link in several of our induction arguments.

The proof of the central limit theorem for $A_n^o(\pi_n^*)$ requires a reasonably detailed understanding of both the threshold functions $g_k(\cdot)$, $1 \leq k < \infty$, and the value functions $v_k(\cdot)$, $1 \leq k < \infty$. Principally, we need to prove all of the monotonicity and convergence properties that are suggested by Figure 1, and this requires a sustained analysis of the Bellman equation (3).

Assuming the few facts reviewed above, the development given here can be read independently of our earlier work. Still, for the purpose of comparison, we should note that the notation used here simplifies our earlier notation in some significant ways. For example, we now take k to be number of observations *yet to be seen*, and this gives us the pleasing formulation (3) of the Bellman equation. We also write $g_k(y)$ for the threshold function when there are k observations yet to be seen, and this replaces the earlier, more cumbersome, notation $f_{n-k+1,n}^*(y)$.

3. GEOMETRY OF THE VALUE AND THRESHOLD FUNCTIONS

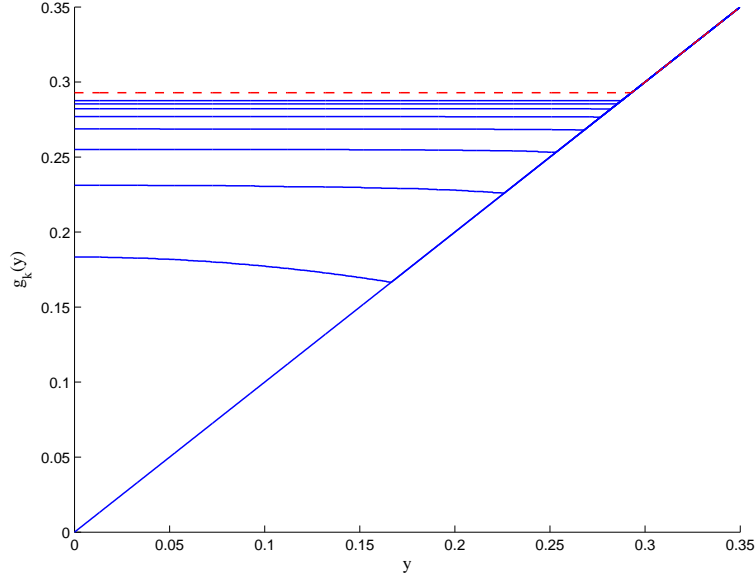
Figure 1 gives a highly suggestive picture of the individual threshold functions $g_k(\cdot)$, and it foretells much of the story about how they behave as $k \rightarrow \infty$. Analytical confirmation of these suggestions is the central challenge. The path to understanding the threshold functions goes through the value functions, and we begin by proving the very plausible fact that the value functions are strictly decreasing.

Lemma 2 (Strict Monotonicity of the Value Functions). *For each $1 \leq k < \infty$, the value function $y \mapsto v_k(y)$ defined by the Bellman recursion (3) is strictly decreasing on $[0, 1]$.*

Proof. The proof uses induction on the sequence of hypotheses:

$$\mathbf{H}_k : \quad v_k(y + \epsilon) < v_k(y) \quad \text{for all } y \in [0, 1) \text{ and all } \epsilon > 0 \text{ such that } y + \epsilon \leq 1.$$

FIGURE 1. The threshold functions g_k , $1 \leq k \leq 10$, (solid lines) and their limit as $k \rightarrow \infty$ (dashed line) for $y \in [0, 35/100]$. The plot suggests most of the analytical properties that are needed for the proof of the central limit theorem.



Since $v_1(y) = 1 - y$, \mathbf{H}_1 is true. For $k \geq 2$, we note by the Bellman recursion (3) that we have

$$\begin{aligned}
 v_k(y + \epsilon) - v_k(y) &= (y + \epsilon)v_{k-1}(y + \epsilon) + \int_{y+\epsilon}^1 \max\{v_{k-1}(y + \epsilon), 1 + v_{k-1}(1 - x)\} dx \\
 &\quad - yv_{k-1}(y) - \int_y^1 \max\{v_{k-1}(y), 1 + v_{k-1}(1 - x)\} dx \\
 &\leq (y + \epsilon)v_{k-1}(y + \epsilon) + \int_{y+\epsilon}^1 \max\{v_{k-1}(y), 1 + v_{k-1}(1 - x)\} dx \\
 &\quad - (y + \epsilon)v_{k-1}(y) - \int_{y+\epsilon}^1 \max\{v_{k-1}(y), 1 + v_{k-1}(1 - x)\} dx \\
 &= (y + \epsilon) \{v_{k-1}(y + \epsilon) - v_{k-1}(y)\} < 0,
 \end{aligned}$$

where the first inequality of the chain follows from

$$\epsilon v_{k-1}(y) \leq \int_y^{y+\epsilon} \max\{v_{k-1}(y), 1 + v_{k-1}(1 - x)\} dx$$

and the second inequality follows from \mathbf{H}_{k-1} . This completes the proof of \mathbf{H}_k and of the lemma. \square

An important benefit of the Bellman recursion (3) is that it provides us with a *variational characterization* of the threshold functions $g_k(\cdot)$, $1 \leq k < \infty$. Specifically, we have the identity

$$(5) \quad g_k(y) = \inf\{x \in [y, 1] : v_{k-1}(y) \leq 1 + v_{k-1}(1-x)\},$$

and, if y is such that $v_{k-1}(y) \leq 1 + v_{k-1}(1-y)$, then $g_k(y)$ is the value for which the decision maker is indifferent between selecting the current observation x (and changing the state of the system from y to $1-x$), or rejecting x (and leaving the state of the system, y , unchanged). By the strict monotonicity of $v_{k-1}(\cdot)$ we then see that $g_k(y)$ is uniquely determined for each $y \in [0, 1]$.

Figure 1 further suggests that the threshold functions have a long interval of fixed points; the next lemma partially confirms this.

Lemma 3 (Range of Fixed Points). *For all $k \geq 1$ and $y \in [0, 1]$ we have*

$$(6) \quad v_k(y) - v_k(2/3) \leq v_k(0) - v_k(2/3) \leq 1.$$

In particular, for all $k \geq 1$ we have

$$(7) \quad g_k(y) = y \quad \text{for all } y \in [1/3, 1]$$

and

$$(8) \quad g_k(y) \leq 1/3 \quad \text{for all } y \in [0, 1/3].$$

Proof. The first inequality of (6) is trivial since the map $y \mapsto v_k(y)$ is strictly decreasing in y . Also, the identities (7) and (8) are immediate from the variational characterization (5) and the bound (6).

The real task is to prove the second inequality of (6). This time we use induction on the hypotheses given by

$$(9) \quad \mathbf{H}_k : \quad v_k(0) - v_k(2/3) \leq 1, \quad \text{for } 1 \leq k < \infty.$$

As before $v_1(y) = 1 - y$, so \mathbf{H}_1 is trivially true. Now, when we apply the Bellman recursion (3) with $y = 0$ and $y = 2/3$ we get

$$\begin{aligned} v_k(0) - v_k(2/3) &= \int_0^1 \max\{v_{k-1}(0), 1 + v_{k-1}(1-u)\} du \\ &\quad - (2/3)v_{k-1}(2/3) - \int_{2/3}^1 \max\{v_{k-1}(2/3), 1 + v_{k-1}(1-u)\} du, \end{aligned}$$

from which a change of variables gives

$$(10) \quad v_k(0) - v_k(2/3) = \int_0^{1/3} I_1(u) du + \int_{1/3}^1 I_2(u) du$$

where $I_1(u)$ and $I_2(u)$ are defined by

$$I_1(u) \equiv \max\{v_{k-1}(0), 1 + v_{k-1}(u)\} - \max\{v_{k-1}(2/3), 1 + v_{k-1}(u)\}$$

and

$$I_2(u) \equiv \max\{v_{k-1}(0) - v_{k-1}(2/3), 1 + v_{k-1}(u) - v_{k-1}(2/3)\}.$$

For the first integrand, $I_1(u)$, we note that

$$(11) \quad \begin{aligned} I_1(u) &= \max\{v_{k-1}(0) - v_{k-1}(2/3), 1 + v_{k-1}(u) - v_{k-1}(2/3)\} \\ &\quad - \max\{0, 1 + v_{k-1}(u) - v_{k-1}(2/3)\}. \end{aligned}$$

The induction assumption \mathbf{H}_{k-1} then tells us that

$$v_{k-1}(0) - v_{k-1}(2/3) \leq 1,$$

and the strict monotonicity of the value function $v_{k-1}(\cdot)$ on $[0, 1]$ yields

$$1 \leq 1 + v_{k-1}(u) - v_{k-1}(2/3) \quad \text{for all } u \in [0, 1/3].$$

Thus, both the first and the second addend in (11) equal the right maximand and

$$(12) \quad I_1(u) = 0 \quad \text{for all } u \in [0, 1/3],$$

so the first integral in (10) vanishes.

To estimate $I_2(u)$ note that \mathbf{H}_{k-1} and monotonicity of $y \mapsto v_{k-1}(y)$ tell us

- if $u \in [1/3, 2/3]$, then

$$I_2(u) = 1 + v_{k-1}(u) - v_{k-1}(2/3) \leq 1 + v_{k-1}(0) - v_{k-1}(2/3) \leq 2 \text{ and}$$

- if $u \in [2/3, 1]$, then

$$I_2(u) = \max\{v_{k-1}(0) - v_{k-1}(2/3), 1 + v_{k-1}(u) - v_{k-1}(2/3)\} \leq 1.$$

Now we just calculate

$$v_k(0) - v_k(2/3) = \int_{1/3}^1 I_2(u) du \leq \int_{1/3}^{2/3} 2 du + \int_{2/3}^1 1 du = 1,$$

and thus we complete the proof of (6). \square

From Lemma 3 we know that a threshold function g_k has many fixed points; in particular, $g_k(y) = y$ if $y \in [1/3, 1]$. Figure 1 further suggests that much of the geometry of g_k is governed by its *minimal* fixed point:

$$(13) \quad \xi_k \equiv \inf\{y : g_k(y) = y\}.$$

The value ξ_k also has a useful policy interpretation. If the value y of the last observation selected is bigger than ξ_k , then the decision maker follows a greedy policy; he accepts *any* feasible arriving observation. On the other hand, if $y < \xi_k$, the decision maker acts conservatively; his choices are governed by the value of the threshold $g_k(y)$. Finally, if $y = \xi_k$, the greedy policy and the optimal policy agree. This interpretation of ξ_k is formalized in the next lemma, where we also prove that the sequence $\{\xi_k : k = 1, 2, \dots\}$ is non-decreasing.

Lemma 4 (Characterization of the Minimal Fixed Point). *For $k \geq 3$, the minimal fixed point $\xi_k \equiv \inf\{y : g_k(y) = y\}$ is the unique solution to the equation*

$$v_{k-1}(y) - v_{k-1}(1 - y) = 1.$$

Moreover, the minimal fixed points form a non-decreasing sequence, so we have

$$(14) \quad \xi_k \leq \xi_{k+1} \quad \text{for all } k \geq 1.$$

Proof. From the variational characterization of $g_k(\cdot)$, we have

$$g_k(y) = \inf\{x \in [y, 1] : v_{k-1}(y) \leq 1 + v_{k-1}(1 - x)\},$$

so if we set $\delta_k(y) \equiv v_{k-1}(y) - v_{k-1}(1 - y)$, then we have

$$(15) \quad g_k(y) = y \quad \text{if and only if } \delta_k(y) \leq 1.$$

The Bellman equation (3) for $v_k(\cdot)$ and Lemma 2 tell us that the map $y \mapsto v_{k-1}(y)$ is continuous and strictly decreasing with $v_1(y) = 1 - y$ and $v_2(y) = (3/2)(1 - y^2)$. Then, the function δ_k is continuous and strictly decreasing, and for $k \geq 3$ we have

$\delta_k(0) = v_{k-1}(0) \geq v_2(0) = 3/2 > 1$, and $\delta_k(1) = -v_{k-1}(0) < 0$, so, there is a unique value y^* such that

$$\delta_k(y^*) \equiv v_{k-1}(y^*) - v_{k-1}(1 - y^*) = 1.$$

Since the map $y \mapsto \delta_k(y)$ is strictly decreasing, we can also write y^* as

$$y^* = \inf\{y : v_{k-1}(y) - v_{k-1}(1 - y) \leq 1\} = \inf\{y : g_k(y) = y\} = \xi_k,$$

where the second equality follows from (15) and the third equality comes from the definition of ξ_k .

To prove the monotonicity property $\xi_k \leq \xi_{k+1}$ for all $k \geq 1$, we first note that since $v_0(y) \equiv 0$ and $v_1(y) \equiv 1 - y$, we have that $\xi_1 = \xi_2 = 0$. Thus, for $k \geq 3$, we find

$$\begin{aligned} \xi_k &= \inf\{y \in [0, 1/3] : g_k(y) = y\} \\ &= \inf\{y \in [0, 1/3] : \delta_k(y) \equiv v_{k-1}(y) - v_{k-1}(1 - y) \leq 1\} \\ (16) \quad &\leq \inf\{y \in [0, 1/3] : \delta_{k+1}(y) \equiv v_k(y) - v_k(1 - y) \leq 1\} \\ &= \inf\{y \in [0, 1/3] : g_{k+1}(y) = y\} = \xi_{k+1}, \end{aligned}$$

where the one inequality (16) follows from restricted supermodularity (4). \square

4. A SECOND SUPERMODULARITY OF THE BELLMAN RECURRENCE

The value functions have a second supermodularity property that provides some crucial help. Specifically, we need it to show that the threshold functions $g_k(\cdot)$ increase with $1 \leq k < \infty$. This monotonicity moves us a long way toward an exhaustive understanding of the asymptotic behavior of the threshold functions.

Proposition 5 (Second Supermodularity). *For all $k \geq 3$, the value functions defined by the Bellman recursion (3) satisfy the bound*

$$(17) \quad v_{k-1}(y) - v_{k-1}(1 - x) \leq v_k(y) - v_k(1 - x) \text{ for all } y \leq \xi_k \text{ and } x \in [y, g_k(y)].$$

Proof. We again use induction to exploit the Bellman equation, and this time the sequence of hypotheses is given by

$$\mathbf{H}_k : v_{k-1}(y) - v_{k-1}(1 - x) \leq v_k(y) - v_k(1 - x), \text{ for all } y \leq \xi_k \text{ and } x \in [y, g_k(y)].$$

We first prove \mathbf{H}_3 , which we then use as the base case for our induction. We recall that $v_1(y) = 1 - y$ and, if we use the Bellman recursion (3), we obtain that $v_2(y) = (3/2)(1 - y^2)$. In turn, this implies $g_3(y) = \max\{1 - \sqrt{2/3 + y^2}, y\}$ and $\xi_3 = 1/6$. To calculate $v_3(y)$ we apply the Bellman recursion one more time, and we obtain a messier but still tractable formula:

$$v_3(y) = \begin{cases} (3/2)(1 - y^2) + 3^{-3/2}(2 + 3y^2)^{3/2} & \text{if } y \leq 1/6 \\ (1/2)(1 - y)(4 + 5y + 2y^2) & \text{if } y \geq 1/6. \end{cases}$$

Thus, for $y \leq \xi_3 = 1/6$, we need to show

$$v_2(y) - v_2(1 - x) \leq v_3(y) - v_3(1 - x) \quad \text{for all } x \in [y, g_3(y)],$$

where $g_3(y) = 1 - \sqrt{2/3 + y^2}$. From our explicit formulas for $v_2(\cdot)$ and $v_3(\cdot)$ we have

$$v_3(1 - x) - v_2(1 - x) = (5/2)x - 3x^2 + x^3,$$

and

$$v_3(y) - v_2(y) = 3^{-3/2}(2 + 3y^2)^{3/2} \geq (2/3)^{3/2} \approx 0.5443.$$

Calculus shows that $(5/2)x - 3x^2 + x^3$ increases on $0 \leq x \leq 1 - \sqrt{2/3}$ and attains an endpoint maximum of $(1/18)(9 - \sqrt{6}) \approx 0.3640$. Thus, we find

$$v_3(1-x) - v_2(1-x) \leq (1/18)(9 - \sqrt{6}) < (2/3)^{3/2} \leq v_3(y) - v_2(y)$$

for all $y \leq 1/6$ and $y \leq x \leq 1 - \sqrt{2/3 + y^2}$, completing the proof of \mathbf{H}_3 .

We now suppose that \mathbf{H}_k holds, and we seek to show \mathbf{H}_{k+1} . First, from the variational characterization of $g_k(\cdot)$ and the definition of ξ_k , recall that

$$1 \leq v_{k-1}(y) - v_{k-1}(1-x) \quad \text{for } y \leq \xi_k \text{ and } x \in [y, g_k(y)],$$

which, together with the induction assumption \mathbf{H}_k , implies

$$(18) \quad 1 \leq v_{k-1}(y) - v_{k-1}(1-x) \leq v_k(y) - v_k(1-x) \quad \text{for } y \leq \xi_k \text{ and } x \in [y, g_k(y)].$$

The second inequality in (18) and the variational characterization (5) give us

$$g_k(y) \leq g_{k+1}(y) \quad \text{for all } y \leq \xi_k.$$

Moreover, if $x \in [g_k(y), g_{k+1}(y)]$ the variational characterization of $g_{k+1}(\cdot)$ also gives

$$v_{k-1}(y) - v_{k-1}(1-x) \leq 1 \leq v_k(y) - v_k(1-x) \quad \text{for } y \leq \xi_k \text{ and } x \in [g_k(y), g_{k+1}(y)],$$

which combines with (18) to give the crucial inequality

$$(19) \quad v_{k-1}(y) - v_{k-1}(1-x) \leq v_k(y) - v_k(1-x) \quad \text{for } y \leq \xi_k \text{ and } x \in [y, g_{k+1}(y)].$$

From an application of the Bellman recursion (3) for $y \leq \xi_k$ and $x \in [y, g_{k+1}(y)]$, we obtain

$$(20) \quad \begin{aligned} v_k(y) - v_k(1-x) &= y(v_{k-1}(y) - v_{k-1}(1-x)) \\ &+ \int_y^{1-x} \max\{v_{k-1}(y) - v_{k-1}(1-x), 1 + v_{k-1}(1-u) - v_{k-1}(1-x)\} du. \end{aligned}$$

If we now change variable in the last integral by replacing u with $1-u$, then the range of integration changes to $[x, 1-y]$ and we can rewrite (20) as

$$\begin{aligned} v_k(y) - v_k(1-x) &= y(v_{k-1}(y) - v_{k-1}(1-x)) \\ &+ \int_x^{1-x} \max\{v_{k-1}(y) - v_{k-1}(1-x), 1 + v_{k-1}(u) - v_{k-1}(1-x)\} du \\ &+ \int_{1-x}^{1-y} \max\{v_{k-1}(y) - v_{k-1}(1-x), 1 + v_{k-1}(u) - v_{k-1}(1-x)\} du. \end{aligned}$$

In this last equation we see that we can use our crucial inequality (19) to bound the first addend and the left maximand of the other two addends. Moreover, since $x \leq g_{k+1}(y) \leq 1/3$, we can appeal to the restricted supermodularity (4) to bound the right maximand of the second addend. In doing so, we obtain

$$(21) \quad \begin{aligned} v_k(y) - v_k(1-x) &\leq y(v_k(y) - v_k(1-x)) \\ &+ \int_x^{1-x} \max\{v_k(y) - v_k(1-x), 1 + v_k(u) - v_k(1-x)\} du \\ &+ \int_{1-x}^{1-y} \max\{v_k(y) - v_k(1-x), 1 + v_{k-1}(u) - v_{k-1}(1-x)\} du. \end{aligned}$$

We now observe that the monotonicity property of the map $u \mapsto v_{k-1}(u)$ for $u \in [1-x, 1-y]$ and the variational characterization of $g_{k+1}(\cdot)$ combine to give

$$1 + v_{k-1}(u) - v_{k-1}(1-x) \leq 1 \leq v_k(y) - v_k(1-x)$$

for all $y \leq \xi_k$ and $x \in [y, g_{k+1}(y)]$. Hence, the third integrand in (21) satisfies the equality

$$\max\{v_k(y) - v_k(1-x), 1 + v_{k-1}(u) - v_{k-1}(1-x)\} = v_k(y) - v_k(1-x),$$

and an analogous monotonicity argument for $u \in [1-x, 1-y]$ also yields

$$\max\{v_k(y) - v_k(1-x), 1 + v_k(u) - v_k(1-x)\} = v_k(y) - v_k(1-x).$$

When we use the last two observations in (21) we obtain that

$$v_k(y) - v_k(1-x) \leq v_{k+1}(y) - v_{k+1}(1-x), \text{ for all } y \leq \xi_k \text{ and } x \in [y, g_{k+1}(y)].$$

We now conclude our argument by considering values $y \in [\xi_k, \xi_{k+1}]$. From the variational characterization of $g_{k+1}(\cdot)$ and the definition of ξ_k , we obtain

$$v_{k-1}(y) - v_{k-1}(1-x) \leq 1 \leq v_k(y) - v_k(1-x) \quad \text{for } y \in [\xi_k, \xi_{k+1}] \text{ and } x \in [y, g_{k+1}(y)]$$

which can be used instead of (19) to construct an argument similar to the earlier one and conclude that

$$v_k(y) - v_k(1-x) \leq v_{k+1}(y) - v_{k+1}(1-x), \text{ for } y \in [\xi_k, \xi_{k+1}] \text{ and } x \in [y, g_{k+1}(y)],$$

just as needed to complete the proof of (17). \square

The usefulness of second supermodularity property Proposition 5 shows itself simply — but clearly — in the following corollary.

Corollary 6 (Monotonicity of Optimal Thresholds). *For all $y \in [0, 1]$ the threshold functions satisfy*

$$(22) \quad g_k(y) \leq g_{k+1}(y) \quad \text{for all } k \geq 1, \text{ and}$$

$$(23) \quad 1/6 \leq g_k(y) \quad \text{for all } k \geq 3.$$

Proof. For $k = 1, 2$ we have $v_0(y) = 0$ and $v_1(y) = 1 - y$, so that

$$g_1(y) = g_2(y) = y.$$

For $k = 3$, we have already noticed in the course of proving Proposition 5 that we have $g_3(y) = \max\{1 - \sqrt{2/3 + y^2}, y\}$, so, in particular, $g_3(y) \geq 1/6$ for $y \in [0, 1]$. Finally, for $k > 3$, the bound (17) and the variational characterization (5) of the threshold function give us (22), and this confirms the lower bound (23). \square

We now pursue two further suggestions from Figure 1. Specifically, we show that the limit function g_∞ has exactly the piecewise linear shape that the figure suggests, and we also show that the convergence to g_∞ is uniform. The proof of these fact requires some additional regularity properties that are discussed in the next section.

5. REGULARITY OF THE VALUE AND THRESHOLD FUNCTIONS

The minimal fixed points give us a powerful guide to the geometry of the value function and its derivatives. The connection begins with the Bellman recursion (3) and the variational characterization (5) which together give the identity

$$v_k(y) = g_k(y)v_{k-1}(y) + \int_{g_k(y)}^1 \{1 + v_{k-1}(1-x)\} dx.$$

If we now differentiate both sides with respect to y , we obtain the recursion for the first derivative:

$$v'_k(y) = g'_k(y)v_{k-1}(y) + g_k(y)v'_{k-1}(y) - g'_k(y) \{1 + v_{k-1}(1 - g_k(y))\}.$$

The definition of the minimal fixed point (13) and the variational characterization (5) then give us

$$(24) \quad v_{k-1}(y) = 1 + v_{k-1}(1 - g_k(y)) \quad \text{if } y \leq \xi_k,$$

so our recursion for $v'_k(\cdot)$ can be written more informatively as

$$(25) \quad v'_k(y) = \begin{cases} g_k(y)v'_{k-1}(y) & \text{if } y \leq \xi_k \\ v_{k-1}(y) - 1 - v_{k-1}(1 - y) + yv'_{k-1}(y) & \text{if } y \geq \xi_k. \end{cases}$$

These relations underscore the importance of the minimal fixed points to the geometry of the value function, and they also lead to useful regularity properties.

Lemma 7 (Monotonicity Properties of the Derivatives). *For all $k \geq 1$, we have*

$$(26) \quad -1 \leq v'_k(y) \leq v'_{k+1}(y) \leq 0 \quad \text{for } y \in [0, 1/3] \text{ and}$$

$$(27) \quad v'_{k+1}(y) \leq v'_k(y) \leq -1 \quad \text{for } y \in [1/2, 1].$$

Proof. We already know from Lemma 2 that $y \mapsto v_k(y)$ is strictly decreasing, so $v'_k(y)$ is non-positive on $[0, 1]$. Since $0 \leq g_k(y) \leq 1$, the top line of (25) tells us that

$$(28) \quad v'_{k-1}(y) \leq g_k(y)v'_{k-1}(y) = v'_k(y) \quad \text{for } y \leq \xi_k.$$

To cover the rest of the range in (26), we use induction on the sequence of hypotheses

$$\mathbf{H}_k : \quad v'_{k-1}(y) \leq v'_k(y), \quad \text{for all } y \in [\xi_k, 1/3] \text{ and } 2 \leq k < \infty.$$

For the base case \mathbf{H}_2 we have $\xi_2 = 0$, $v_1(y) = 1 - y$, and $v_2(y) = (3/2)(1 - y^2)$. So

$$v'_1(y) = -1 \leq -3y = v'_2(y) \quad \text{if and only if } y \leq 1/3,$$

just as needed. Now taking \mathbf{H}_k as our induction assumption, we seek to prove \mathbf{H}_{k+1} .

First, for $y \in [\xi_k, 1/3]$, the second line of (25) gives us $v'_k(\cdot)$. By restricted supermodularity (4), the monotonicity $\xi_k \leq \xi_{k+1}$, and the induction assumption \mathbf{H}_k , we see for $y \in [\xi_{k+1}, 1/3]$ that

$$\begin{aligned} v'_k(y) &= v_{k-1}(y) - 1 - v_{k-1}(1 - y) + yv'_{k-1}(y) \\ &\leq v_k(y) - 1 - v_k(1 - y) + yv'_k(y) = v'_{k+1}(y), \end{aligned}$$

completing the proof \mathbf{H}_{k+1} . To complete the proof of (26), one just needs to note that the lower bound $-1 \leq v'_k(y)$ now follows from $v'_1(y) = -1$ together with (28) and \mathbf{H}_k .

To prove (27), we again use induction, but this time the sequence of hypothesis is given by

$$\mathbf{H}_k : \quad v'_k(y) \leq v'_{k-1}(y) \quad \text{for } y \in [1/2, 1], \text{ and } 2 \leq k < \infty.$$

As before, $v_1(y) = 1 - y$ and $v_2(y) = (3/2)(1 - y^2)$ so $v'_1(y) = -1$ and $v'_2(y) = -3y$. For $y \geq 1/2$, we then have

$$v'_2(y) \leq -3/2 \leq -1 = v'_1(y),$$

proving \mathbf{H}_2 . As tradition demands, we again take \mathbf{H}_k as our induction assumption, and we seek to prove \mathbf{H}_{k+1} .

Since $y \in [1/2, 1]$ we have $1 - y \leq 1/2 \leq y$, so the restricted supermodularity property (4) gives us

$$(29) \quad v_{k-1}(1 - y) - v_{k-1}(y) \leq v_k(1 - y) - v_k(y).$$

Next, recall the identity of the bottom line of (25), but, as you do so, replace k by $k + 1$. We can then directly apply (29) and \mathbf{H}_k to get

$$\begin{aligned} v'_{k+1}(y) &= v_k(y) - 1 - v_k(1 - y) + yv'_k(y) \\ &\leq v_{k-1}(y) - 1 - v_{k-1}(1 - y) + yv'_{k-1}(y) = v'_k(y). \end{aligned}$$

This inequality completes the proof of \mathbf{H}_{k+1} and confirms the lower bound of (27). For the upper bound of (27), $v'_k(y) \leq -1$ on $[1/2, 1]$, we just need to note that it follows from the fact $v'_1(y) = -1$ and the validity of \mathbf{H}_k for all $k \geq 1$. \square

The smoothness of the value functions converts easily into a very useful Lipschitz equi-continuity property of the threshold functions.

Lemma 8 (Lipschitz Equi-Continuity of Threshold Functions). *For all $k \geq 1$ we have*

$$(30) \quad |g_k(y) - g_k(z)| \leq |y - z| \quad \text{for all } y, z \in [0, 1].$$

Proof. We first consider $y \in [0, \xi_k]$. In this case, we have that identity (24) holds, by its differentiation, we obtain

$$(31) \quad g'_k(y) = -\frac{|v'_{k-1}(y)|}{|v'_{k-1}(1 - g_k(y))|} \leq 0 \quad \text{for all } y \in [0, \xi_k].$$

Moreover, since $y \in [0, \xi_k]$ we know that $y \leq 1/3$ so by (8) we have $g_k(y) \leq 1/3$, and hence by (27) we obtain $1 \leq |v'_{k-1}(1 - g_k(y))|$. Consequently, (31) gives us

$$(32) \quad |g'_k(y)| \leq |v'_{k-1}(y)| \quad \text{for all } y \in [0, \xi_k],$$

and (26) implies $|v'_k(y)| \leq 1$. Thus, at last, we have the uniform bound

$$(33) \quad |g'_k(y)| \leq 1 \quad \text{for all } y \in [0, \xi_k],$$

which confirms the inequality (30) for $y, z \in [0, \xi_k]$. Also, for $y, z \in [\xi_k, 1]$ we have that (30) trivially holds, so if we choose $y < \xi_k < z$, the triangle inequality gives us

$$|g_k(y) - g_k(z)| \leq |g_k(y) - g_k(\xi_k)| + |g_k(\xi_k) - g_k(z)| \leq |y - z|,$$

confirming that (30) holds in general. \square

6. THE OPTIMAL POLICY AT INFINITY

The minimal fixed points ξ_k , $k = 1, 2, \dots$, are non-decreasing and bounded by $1/3$, so they have a limit

$$(34) \quad \lim_{k \rightarrow \infty} \xi_k \stackrel{\text{def}}{=} \xi \leq 1/3.$$

The threshold values $g_k(y)$, $k = 1, 2, \dots$ are also non-decreasing and bounded, so they have a pointwise limit $g_\infty(y)$. The next lemma characterizes g_∞ and gives a crucial bound on the uniform rate of convergence to g_∞

Proposition 9 (Characterization of Limiting Threshold). *For the limit threshold g_∞ , we have the formula*

$$g_\infty(y) = \max\{\xi, y\} \quad \text{for all } y \in [0, 1].$$

Moreover, we have an exact measure of the uniform rate of convergence

$$(35) \quad \max_{0 \leq y \leq 1} |g_k(y) - g_\infty(y)| = \xi - \xi_k \quad \text{for all } k \geq 1.$$

Proof. We first fix m and $y \in [0, \xi_m]$. We then recall that $y \leq \xi_m \leq 1/3$ implies that $g_j(y) \leq 1/3$ for all $j \geq 1$. Now, given $k \geq m$ we can repeatedly apply the top line of (25) to obtain

$$(36) \quad |v'_k(y)| = |v'_{m-1}(y)| \left(\prod_{j=m}^k g_j(y) \right) \leq 3^{m-k} |v'_{m-1}(y)| \quad \text{for } y \in [0, \xi_m],$$

and by (26) we have $|v'_{m-1}(y)| \leq 1$ for all $y \in [0, 1/3]$, so (32) gives us more simply

$$(37) \quad \max_{0 \leq y \leq \xi_m} |g'_k(y)| \leq 3^{m-k} \quad \text{for all } k \geq m.$$

Now, for any y, z in $[0, \xi_m]$ we have $|g_k(y) - g_k(z)| \leq 3^{m-k} |y - z|$ so, letting $k \rightarrow \infty$, we obtain that g_∞ is constant on $[0, \xi_m]$ for each $m \geq 1$. Since $\xi_m \uparrow \xi$, there is a constant c such that $g_\infty(y) = c$ for all $y \in [0, \xi]$.

As Figure 1 suggests, $c = \xi$ and this is easy to confirm. Again we fix m , take $k \geq m$, and note that by the triangle inequality and the Lipschitz bound (30) on g_k we have

$$\begin{aligned} |g_\infty(\xi_m) - \xi_k| &\leq |g_\infty(\xi_m) - g_k(\xi_m)| + |g_k(\xi_m) - g_k(\xi_k)| \\ &\leq |g_\infty(\xi_m) - g_k(\xi_m)| + |\xi_m - \xi_k|. \end{aligned}$$

When $k \rightarrow \infty$, $g_k(\xi_m)$ converges to $g_\infty(\xi_m)$ and ξ_k to ξ so we have

$$|g_\infty(\xi_m) - \xi| \leq |\xi_m - \xi|.$$

Since $g_\infty(\xi_m) = c$ does not depend on m and since $|\xi_m - \xi| \rightarrow 0$ as $m \rightarrow \infty$, we see that $g_\infty(\xi_m) = \xi$ for all $m \geq 1$ and consequently $g_\infty(y) = \xi$ for all $y \in [0, \xi]$. Finally, for all $m \geq 1$ we also have $g_m(y) = y$ for each $y \in [\xi, 1]$, so the proof of the formula for g_∞ is complete.

To prove (35), we first note

$$g_\infty(y) - g_k(y) = \begin{cases} \xi - g_k(y) & y \in [0, \xi_k], \\ \xi - y & y \in [\xi_k, \xi], \\ 0 & y \in [\xi, 1]. \end{cases}$$

By (31), $g_k(y)$ is strictly decreasing on $[0, \xi_k]$, so the gap $g_\infty(y) - g_k(y)$ is maximized when $y = \xi_k$. This gap decreases linearly over the interval $[\xi_k, \xi]$ and equals 0 at ξ ; consequently the maximal gap is exactly equal to $\xi - \xi_k$. \square

7. THE CENTRAL LIMIT THEOREM FOR $A_n^o(\pi_\infty)$ IS EASY

We now recall that ξ denotes the limit (34) of the minimal fixed points, and we define a selection policy π_∞ for all X_1, X_2, \dots by taking the (time independent) threshold function to be

$$g_\infty(y) = \max\{\xi, y\} \equiv \xi \vee y.$$

If $A_n^o(\pi_\infty)$ counts the number of selections made by policy π_∞ up to and including time n , then we have the explicit formula

$$(38) \quad A_n^o(\pi_\infty) = \sum_{i=1}^n \mathbb{1}(X_i \geq \xi \vee Y'_{i-1}),$$

where one sets $Y'_0 = 0$, and one defines Y'_i for $i \geq 1$ recursively by

$$(39) \quad Y'_i = \begin{cases} Y'_{i-1} & \text{if } X_i < \xi \vee Y'_{i-1} \\ 1 - X_i & \text{if } X_i \geq \xi \vee Y'_{i-1}. \end{cases}$$

Given the facts that have been accumulated, it turns out to be a reasonably easy task to prove a central limit theorem for $A_n^o(\pi_\infty)$. One just needs to make the right connection to the known central limit theorems for Markov additive processes.

To make this connection explicit, we first consider the bi-variate random sequence $\{Z_i = (X_i, Y'_{i-1}) : i = 1, 2, 3, \dots\}$ and note that it may be viewed as a two-dimensional Markov chain on the state space $\mathcal{S} \equiv [0, 1] \times [0, 1 - \xi]$. Specifically, for any pair $(x, y) \in \mathcal{S}$ and any Borel set $C \subseteq \mathcal{S}$, we have the point-to-set transition kernel

$$K((x, y), C) = \int_0^1 [\mathbb{1}\{(u, 1 - x) \in C\} \mathbb{1}(x \geq \xi \vee y) + \mathbb{1}\{(u, y) \in C\} \mathbb{1}(x < \xi \vee y)] du.$$

Given this explicit formula, it is a straightforward (but admittedly a little tedious) to check that a stationary probability measure for the kernel K is given by the uniform distribution γ on $\mathcal{S} = [0, 1] \times [0, 1 - \xi]$. We will confirm shortly that γ is also the unique stationary distribution.

To more deeply understand the chain Z_i , $i = 1, 2, \dots$, we now consider the double chain (Z_i, \bar{Z}_i) , $i = 1, 2, \dots$, where $Z_1 = (x, y)$ is an arbitrary point of \mathcal{S} and \bar{Z}_1 has the uniform distribution on \mathcal{S} . For $i = 1, 2, \dots$, the chains $\{Z_i = (X_i, Y'_{i-1})\}$ and $\{\bar{Z}_i = (X_i, \bar{Y}'_{i-1})\}$ share the same independent uniform sequence X_i , $i = 1, 2, \dots$, as their first coordinate, while their second coordinates Y'_{i-1} and \bar{Y}'_{i-1} are both determined by the recursions (39). Typically these coordinates differ because of their differing initial values, but we will check that they do not differ for long.

To make this precise, we set $\nu = \min\{i \geq 1 : X_i \geq 1 - \xi\}$, and we show that ν is a *coupling time* for (Z_i, \bar{Z}_i) in the sense that

$$Z_i = \bar{Z}_i \quad \text{for all } i > \nu.$$

Since Y'_i and \bar{Y}'_i both satisfy the recursion (39), we have

$$Y'_i \leq 1 - \xi \quad \text{and} \quad \bar{Y}'_i \leq 1 - \xi \quad \text{for all } i = 1, 2, \dots,$$

so by the definition of ν , we must have

$$\max\{\xi \vee Y'_{\nu-1}, \xi \vee \bar{Y}'_{\nu-1}\} \leq X_\nu.$$

The recursion (39) then gives us

$$Y'_\nu = \bar{Y}'_\nu = 1 - X_\nu \quad \text{and} \quad Z_\nu = \bar{Z}_\nu.$$

By the construction of the double process (Z_i, \bar{Z}_i) , if one has $Z_i(\omega) = \bar{Z}_i(\omega)$ for some $i = i(\omega)$, then $Z_j(\omega) = \bar{Z}_j(\omega)$ for all $j \geq i(\omega)$, so ν is indeed a coupling time for (Z_i, \bar{Z}_i) .

The coupling inequality (see, e.g., Lindvall, 2002, p. 12) then tells us that for all Borel sets $C \subseteq \mathcal{S}$ we have

$$(40) \quad \|K^\ell((x, y), C) - \gamma(C)\|_{\text{TV}} \leq \mathbb{P}(\nu > \ell) = (1 - \xi)^\ell,$$

where γ is the uniform stationary distribution on \mathcal{S} . The bound (40) has several useful implications. First, it implies that γ is the *unique* stationary distribution for the chain with kernel K . It also implies (see, e.g., Meyn and Tweedie, 2009, Theorem 16.0.1) that the chain $\{Z_i : i = 1, 2, \dots\}$ is uniformly ergodic; more specifically, it is a ϕ -mixing chain with

$$\phi(\ell) \leq 2\rho^\ell \quad \text{and} \quad \rho = 1 - \xi.$$

If we set $z = (x, y)$ and $f(z) = \mathbb{1}(x \geq y \vee \xi)$, then the representation (38) can also be written in terms the chain $\{Z_i : i = 1, 2, \dots\}$ as

$$A_n^o(\pi_\infty) = \sum_{i=1}^n f(Z_i),$$

and this makes it explicit that $A_n^o(\pi_\infty)$ is a Markov additive process. Our coupling and the uniform ergodicity of $\{Z_i : i = 1, 2, \dots\}$ imply (see, e.g., Meyn and Tweedie, 2009, Theorem 17.5.3 and Lemma 17.5.1) that there is a constant $\sigma^2 \geq 0$ such that

$$(41) \quad \lim_{n \rightarrow \infty} n^{-1} \text{Var}(A_n^o(\pi_\infty)) = \lim_{n \rightarrow \infty} n^{-1} \text{Var}_\gamma(A_n^o(\pi_\infty)) = \sigma^2,$$

where the first variance refers to the chain started at $Z_1 = (X_1, 0)$ and the second variance refers to the chain started at Z_1 with the stationary distribution γ (i.e. the uniform distribution on \mathcal{S}). The general theory also provides the series representation for the limit (41):

$$(42) \quad \begin{aligned} \sigma^2 &= \text{Var}_\gamma [\mathbb{1}(X_1 \geq \{\xi \vee Y'_0\})] \\ &+ 2 \sum_{i=2}^{\infty} \text{Cov}_\gamma [\mathbb{1}(X_1 \geq \{\xi \vee Y'_0\}), \mathbb{1}(X_i \geq \{\xi \vee Y'_{i-1}\})], \end{aligned}$$

where γ again refers to the situation in which the chain starts with Z_1 having the stationary distribution.

The general representations (41) and (42) give us the existence of σ but they do not automatically entail $\sigma^2 > 0$, so to prove a central limit theorem for $A_n^o(\pi_\infty)$ with the classical normalization, one must independently establish that $\sigma^2 > 0$. To show this, we first need an elementary lemma that provides a variance analog to the information processing inequality for entropy.

Lemma 10 (Information Processing Lemma). *If a random variable X has values in $\{1, 2, \dots\}$ and $P(X = 1) = p$, then $p(1 - p) \leq \text{Var}(X)$.*

Proof. Define a function f on the natural numbers \mathbb{N} by setting $f(1) = 0$ and $f(k) = 1$ for $k \geq 1$. We then have $|f(x) - f(y)| \leq |x - y|$ for all $x, y \in \mathbb{N}$. If we take Y to be an independent copy of X , then we have

$$2p(1-p) = E[(f(X) - f(Y))^2] \leq E[(X - Y)^2] = 2 \text{Var}(X).$$

□

Now we can address the main lemma of this section.

Lemma 11. *The variance of $A_n^o(\pi_\infty)$ satisfies the asymptotic lower bound*

$$\text{Var}(A_n^o(\pi_\infty)) = \Omega(n), \quad \text{as } n \rightarrow \infty.$$

Proof. We first set $\nu_0 \equiv 0$ and then we define the stopping times

$$\nu_t = \inf\{i > \nu_{t-1} : X_i \geq 1 - \xi\}, \quad t = 1, 2, \dots$$

We also set $T = \inf\{t : \nu_t \geq n\}$, and note that T is a stopping time with respect to the increasing sequence of σ -fields

$$\mathcal{G}_t = \sigma\{\nu_1, \nu_2, \dots, \nu_t\} \quad \text{for all } t \geq 1.$$

Next, we set

$$(43) \quad U_t = \sum_{i=\nu_{t-1}+1}^{\nu_t} \mathbb{1}(X_i \geq \xi \vee Y'_{i-1}) \quad \text{for } 1 \leq t \leq T \quad \text{and set}$$

$$V = \sum_{i=n+1}^{\nu_T} \mathbb{1}(X_i \geq \xi \vee Y'_{i-1}),$$

so we have the representation

$$(44) \quad A_n^o(\pi_\infty) = A_{\nu_T}^o(\pi_\infty) - V = \sum_{t=1}^T U_t - V.$$

Here, the random variables U_t , $t = 1, 2, \dots$, are independent and identically distributed. We also have $V \leq \nu_T - n$ and $\nu_T = \inf\{i \geq n : X_i \geq 1 - \xi\}$, so the variance of V is bounded by a constant that depends only on ξ . The existence of the limit (41) and the Cauchy-Schwarz inequality then give us

$$(45) \quad \text{Var}(A_n^o(\pi_\infty)) = \text{Var}(A_{\nu_T}^o(\pi_\infty)) + O(\sqrt{n}) \quad \text{as } n \rightarrow \infty,$$

so to prove the lemma it suffices to show $\text{Var}(A_{\nu_T}^o(\pi_\infty)) = \Omega(n)$.

By the definition of ν_T and U_t , $t = 1, 2, \dots$, we have

$$A_{\nu_T}^o(\pi_\infty) = \sum_{t=1}^T U_t$$

so, by the conditional variance formula, the independence of the U_t 's, and fact that T is \mathcal{G}_T measurable, we have the bound

$$(46) \quad \text{Var}\left(\sum_{t=1}^T U_t\right) \geq \mathbb{E}\left[\text{Var}\left(\sum_{t=1}^T U_t \mid \mathcal{G}_T\right)\right] = \mathbb{E}\left[\sum_{t=1}^T \text{Var}(U_t \mid \mathcal{G}_T)\right].$$

We now note from the definition (43) that U_t takes values in $\{1, 2, \dots, \nu_t - \nu_{t-1}\}$. Thus, if p is the probability that no X_i is selected for $i \in \{\nu_{t-1} + 1, \dots, \nu_t - 1\}$, then setting $a = (1 - \xi)^{-1}\xi$, we have

$$p = \mathbb{P}(U_t = 1 \mid \mathcal{G}_T) = \mathbb{P}(X_i < \xi \text{ for all } \nu_{t-1} + 1 \leq i \leq \nu_t - 1 \mid \mathcal{G}_T) = a^{\nu_t - \nu_{t-1} - 1}.$$

Now, by applying Lemma 10 to the conditional expectation, we have

$$\text{Var}(U_t | \mathcal{G}_T) \geq a^{\nu_t - \nu_{t-1} - 1} (1 - a^{\nu_t - \nu_{t-1} - 1}),$$

so from (46) we have

$$\text{Var}\left(\sum_{t=1}^T U_t\right) \geq \mathbb{E}\left[\sum_{t=1}^T a^{\nu_t - \nu_{t-1} - 1} (1 - a^{\nu_t - \nu_{t-1} - 1})\right].$$

The summands are independent and identically distributed and T is a stopping time with respect to the increasing sequence of σ -fields $\mathcal{G}_t = \sigma\{\nu_1, \nu_2, \dots, \nu_t\}$, $t \geq 1$, so by Wald's identity we have

$$(47) \quad \text{Var}\left(\sum_{t=1}^T U_t\right) \geq \mathbb{E}[T] \mathbb{E}[a^{\nu_1 - 1} (1 - a^{\nu_1 - 1})].$$

For the stopping time T , we have the alternative representation

$$T = 1 + \sum_{i=1}^{n-1} \mathbb{1}(X_i \geq 1 - \xi),$$

so we have $\mathbb{E}[T] = \xi n + O(1)$. Since ν_1 has the geometric distribution with success probability ξ we also have $\mathbb{E}[a^{\nu_1 - 1} (1 - a^{\nu_1 - 1})] > 0$, so by (45) and (47) the proof of the lemma is complete. \square

All of the pieces are now in place. By the central limit theorem for functions of uniformly ergodic Markov chains (Meyn and Tweedie, 2009, Theorem 17.5.3; or Jones, 2004, Corollary 5) we get our central limit theorem for $A_n^o(\pi_\infty)$.

Proposition 12 (Central Limit Theorem for $A_n^o(\pi_\infty)$). *As $n \rightarrow \infty$, we have the limit*

$$\frac{A_n^o(\pi_\infty) - \mu n}{\sqrt{n}} \implies N(0, \sigma^2),$$

where $\mu = \mathbb{E}_\gamma[\mathbb{1}(X_1 \geq \{\xi \vee Y'_0\})]$, γ is the stationary distribution for the Markov chain $\{Z_i : i = 1, 2, \dots\}$, and σ^2 is the constant defined by either the limits (41) or the sum (42).

By appealing to the known relation (1) that $\mathbb{E}[A_n^o(\pi_n^*)] = (2 - \sqrt{2})n + O(1)$, one can show with a bit of calculation that here we have $\mu = 2 - \sqrt{2}$. Since this identification is implicit in the calculations of the next section, there is no reason to belabor it here.

8. $A_n^o(\pi_n^*)$ AND $A_n^o(\pi_\infty)$ ARE CLOSE IN L^2

Proposition 12 tells us that the easy sum $A_n^o(\pi_\infty)$ obeys a central limit theorem, and now the task is to show that the harder sum $A_n^o(\pi_n^*)$ follows the same law. The essence is to show that, after centering, the random variables $A_n^o(\pi_n^*)$ and $A_n^o(\pi_\infty)$ are close in L^2 in the sense that $\|A_n^o(\pi_n^*) - A_n^o(\pi_\infty) - \mathbb{E}[A_n^o(\pi_n^*) - A_n^o(\pi_\infty)]\|_2 = o(\sqrt{n})$ as $n \rightarrow \infty$. For technical convenience, we work with the random variable

$$\Delta_n \stackrel{\text{def}}{=} A_{n-2}^o(\pi_n^*) - \mathbb{E}[A_{n-2}^o(\pi_n^*)] - A_{n-2}^o(\pi_\infty) + \mathbb{E}[A_{n-2}^o(\pi_\infty)].$$

The essential estimate of our development is given by the next lemma. In one way or another, the proof of the lemma calls on all of the machinery that has been developed.

Lemma 13 (L^2 -Estimate). *There is a constant C such that for all $n \geq 3$ we have*

$$\|\Delta_n\|_2^2 \leq C \sum_{k=3}^n (\xi - \xi_k);$$

so, in particular, we have the asymptotic estimate

$$\|\Delta_n\|_2 = o(\sqrt{n}) \quad \text{as } n \rightarrow \infty.$$

Proof. We first note that the threshold function lower bound (23) implies that $Y_i \leq 5/6$ for all $1 \leq i \leq n-2$. Consequently, if $X_i \geq 5/6$, then X_i is selected by both of the policies π_n^* and π_∞ . At such a time i , we have a kind of “renewal event,” though we still have to be attentive to the non-homogeneity of the selection process driven by π_n^* .

To formalize this notion, we set $\tau_0 = 0$ and for $m \geq 1$ we define stopping times

$$\tau_m = \inf \{i > \tau_{m-1} : X_i \geq 5/6\} \quad \text{and} \quad \tau'_m = \min\{\tau_m, n-2\};$$

so τ_m is the time at which the m th “renewal” is observed. For each $1 \leq j \leq n-2$, we then set

$$N(j) = \sum_{i=1}^j \mathbb{1}(X_i \geq 5/6),$$

so the time $\tau_{N(j)}$ is the time of the last renewal up to or equal to j , the time $\tau_{N(j)+1}$ is the time of the first renewal strictly after j , and we have the inclusion

$$\tau_{N(j)} \leq j < \tau_{N(j)+1}.$$

For $1 \leq j \leq n-2$, we then consider the martingale differences defined by

$$d_j = \mathbb{E}[A_{n-2}^o(\pi_n^*) - A_{n-2}^o(\pi_\infty) | \mathcal{F}_j] - \mathbb{E}[A_{n-2}^o(\pi_n^*) - A_{n-2}^o(\pi_\infty) | \mathcal{F}_{j-1}],$$

where \mathcal{F}_0 is the trivial σ -field and $\mathcal{F}_j = \sigma\{X_1, X_2, \dots, X_j\}$ for $1 \leq j \leq n$. Using the counting variables

$$\eta_i \equiv \mathbb{1}(X_i \geq g_{n-i+1}(Y_{i-1})) \quad \text{and} \quad \eta'_i \equiv \mathbb{1}(X_i \geq \xi \vee Y'_{i-1}),$$

we have the tautology

$$\begin{aligned} (48) \quad d_j &= \mathbb{E}\left[\sum_{i=j}^{\tau'_{N(j)+1}} (\eta_i - \eta'_i) | \mathcal{F}_j\right] - \mathbb{E}\left[\sum_{i=j}^{\tau'_{N(j)+1}} (\eta_i - \eta'_i) | \mathcal{F}_{j-1}\right] \\ &\quad + \mathbb{E}\left[\sum_{i=\tau'_{N(j)+1}+1}^{n-2} (\eta_i - \eta'_i) | \mathcal{F}_j\right] - \mathbb{E}\left[\sum_{i=\tau'_{N(j)+1}+1}^{n-2} (\eta_i - \eta'_i) | \mathcal{F}_{j-1}\right], \end{aligned}$$

and this becomes more interesting after one checks that the last two terms cancel.

To confirm the cancelation, we first recall that for $\tau_{N(j)+1} < n-2$ the value $X_{\tau_{N(j)+1}} \geq 5/6$ is selected as member of the alternating subsequence under both policies π_n^* and π_∞ , so we also have

$$Y_{\tau_{N(j)+1}} = Y'_{\tau_{N(j)+1}} = 1 - X_{\tau_{N(j)+1}}.$$

Any difference in the selections that are made by the policies π_n^* and π_∞ after time $\tau_{N(j)+1}$ is measurable with respect to the σ -field

$$\mathcal{T}_j = \sigma\{X_{\tau_{N(j)+1}}, X_{\tau_{N(j)+1}+1}, \dots, X_{n-2}\}.$$

Trivially, we have $j < \tau_{N(j)+1}$, so \mathcal{F}_j is independent of \mathcal{T}_j , and the last two addends in (48) do cancel as claimed.

We can therefore write

$$(49) \quad d_j = \mathbb{E} \left[\sum_{i=j}^{\tau'_{N(j)+1}} (\eta_i - \eta'_i) \mid \mathcal{F}_j \right] - \mathbb{E} \left[\sum_{i=j}^{\tau'_{N(j)+1}} (\eta_i - \eta'_i) \mid \mathcal{F}_{j-1} \right] = W_j - I_{j-1}(W_j),$$

where W_j denotes the first summand and I_{j-1} is the projection onto $L^2(\mathcal{F}_{j-1})$. Denoting the identity by I , we have that $I - I_{j-1}$ is an L^2 contraction, so

$$(50) \quad \mathbb{E} [d_j^2] \leq \mathbb{E} [W_j^2] = \mathbb{E} \left[\left(\sum_{i=j}^{\tau'_{N(j)+1}} (\eta_i - \eta'_i) \right)^2 \right],$$

and the remaining task is to estimate the last right-hand side.

For $1 \leq j \leq n-2$, we let $L(j)$ denote time from j since the last renewal preceding j ; in other words, $L(j)$ is the *age* at time j . Analogously, we let $M(j)$ denote the time from j until the time of the next renewal or until time $n-2$; so $M(j)$ is the *residual life* at time j with truncation at time $n-2$. We then have

$$L(j) = j - \tau_{N(j)} \quad \text{and} \quad M(j) = \tau'_{N(j)+1} - j.$$

Our interarrival times are geometric, so $L(j)$ and $M(j)$ are independent, and for $p = 1/6$ we have

$$\mathbb{P}(L(j) = \ell) = \begin{cases} p(1-p)^\ell & \text{if } 0 \leq \ell < j \\ (1-p)^j & \text{if } \ell = j, \end{cases}$$

and

$$\mathbb{P}(M(j) = m) = \begin{cases} p(1-p)^{m-1} & \text{if } 1 \leq m < n-2-j \\ (1-p)^{n-3-j} & \text{if } m = n-2-j. \end{cases}$$

We now introduce the *disagreement* set

$$D_j[\ell, m] = \{\omega : \exists i \in \{j-\ell+1, \dots, j, \dots, j+m\} : X_i(\omega) \in [\xi_{n-i+1}, \xi]\};$$

this is precisely the set of ω for which, if $Y_{j-\ell} = Y'_{j-\ell}$, then the policies π_∞ and π_n^* differ in at least one selection during the time interval $\{j-\ell+1, \dots, j+m\}$, while on the complementary set $D_j^c[\ell, m]$ the selections all agree. Thus, by the crudest possible bound, we have

$$\left| \sum_{i=j}^{\tau'_{N(j)+1}} (\eta_i - \eta'_i) \right| \leq (L(j) + M(j)) \mathbb{1}(D_j[L(j), M(j)]),$$

and when we square both sides and rearrange, we obtain

$$(51) \quad \begin{aligned} \left(\sum_{i=j}^{\tau'_{N(j)+1}} (\eta_i - \eta'_i) \right)^2 &\leq (L(j) + M(j))^2 \mathbb{1}(D_j[L(j), M(j)]) \\ &= \sum_{m=1}^{n-2-j} \sum_{\ell=0}^j (\ell + m)^2 \mathbb{1}(D_j[\ell, m]) \mathbb{1}(L(j) = \ell) \mathbb{1}(M(j) = m). \end{aligned}$$

For each $1 \leq j \leq n-2$ we now set

$$R_j[\ell, m] = \{\omega : X_i(\omega) < 5/6 \text{ for all } i \in \{j-\ell+1, \dots, j+m\}\},$$

so, $R_j[\ell, m]$ is the event that no renewal takes place in $[j - \ell + 1, j]$ or in $[j + 1, j + m]$. By the definition of $L(j)$ and $M(j)$ we then have

$$\mathbb{1}(L(j) = \ell) = \mathbb{1}(R_j[\ell, 0]) \mathbb{1}(X_{j-\ell} \geq 5/6 \text{ or } \ell = j), \quad \text{for } 0 \leq \ell \leq j,$$

and

$$\mathbb{1}(M(j) = m) \leq \mathbb{1}(R_j[0, m - 1]), \quad \text{for } 1 \leq m \leq n - 2 - j.$$

Thus, if we define $\mathbb{1}(R_j[0, 0]) \equiv 1$, then we have the composite bound

$$(52) \quad \mathbb{1}(L(j) = \ell) \mathbb{1}(M(j) = m) \leq \mathbb{1}(R_j[\ell, m - 1]) \mathbb{1}(X_{j-\ell} \geq 5/6 \text{ or } \ell = j),$$

so by inserting (52) in (51) and recalling (50), we find

$$(53) \quad \mathbb{E}[d_j^2] \leq \sum_{m=1}^{n-2-j} \sum_{\ell=0}^j (\ell + m)^2 \mathbb{E}[\mathbb{1}(D_j[\ell, m]) \mathbb{1}(R_j[\ell, m - 1]) \mathbb{1}(X_{j-\ell} \geq 5/6 \text{ or } \ell = j)].$$

The expected value on the right-hand side of (53) accounts for the probability that policies π_n^* and π_∞ differ when one renewal has occurred at time $j - \ell$, and no renewal will occur until time $j + m$. For this to happen, we need at least one $i \in \{j - \ell + 1, \dots, j + m\}$ such that $X_i \in [\xi_{n-i+1}, \xi]$. Since the X_i 's are uniformly distributed on $[0, 1]$, the probability that $X_i \in [\xi_{n-i+1}, \xi]$ equals $\xi - \xi_{n-i+1}$ and, by the monotonicity of the minimal fixed points in Lemma 4, we have the upper bound $\xi - \xi_{n-i+1} \leq \xi - \xi_{n-(j+m)+1}$ for all $i \in \{j - \ell + 1, \dots, j + m\}$. Then, we can estimate the right-hand side of (53) with Boole's inequality, and obtain that there is a constant C such that

$$\mathbb{E}[\mathbb{1}(D_j[\ell, m]) \mathbb{1}(R_j[\ell, m - 1]) \mathbb{1}(X_{j-\ell} \geq 5/6 \text{ or } \ell = j)] \leq C(m - \ell)(\xi - \xi_{n-(j+m)+1}).$$

At this point, $C = 6/5$ would suffice, but subsequently C denotes a Hardy-style constant that may change from line to line. If we use this last bound in (53), we obtain

$$\mathbb{E}[d_j^2] \leq C \sum_{m=1}^{n-2-j} \sum_{\ell=0}^j (\ell + m)^3 (\xi - \xi_{n-(j+m)+1}) (1 - p)^{\ell+m-1},$$

so, if we change variable by applying the transformation $r = j + m$, we have

$$\mathbb{E}[d_j^2] \leq C \sum_{r=j+1}^{n-2} \sum_{\ell=0}^j (\ell + r - j)^3 (\xi - \xi_{n-r+1}) (1 - p)^{\ell+r-j-1}.$$

If we now sum over $1 \leq j \leq n - 2$, we obtain

$$\mathbb{E}[\Delta_n^2] = \sum_{j=1}^{n-2} \mathbb{E}[d_j^2] \leq C \sum_{j=1}^{n-2} \sum_{r=j+1}^{n-2} \sum_{\ell=0}^j (\ell + r - j)^3 (\xi - \xi_{n-r+1}) (1 - p)^{\ell+r-j-1},$$

so if we interchange the first with the second sum and rearrange, we have

$$\mathbb{E}[\Delta_n^2] \leq C \sum_{r=2}^{n-2} (\xi - \xi_{n-r+1}) \left\{ \sum_{j=1}^{r-1} \sum_{\ell=0}^j (\ell + r - j)^3 (1 - p)^{\ell+r-j-1} \right\}.$$

At this point, it is elementary to check that for all r the last double sum is bounded by the constant $\sum_{u=1}^{\infty} u^4 (1 - p)^{u-1}$, and this completes the proof of our lemma. \square

9. SOME PERSPECTIVE

We have pursued the proof of a specific central limit theorem, but our analysis may be viewed more generally as a case study for a substantial class of Markov decision problems (MDPs). Here, we took advantage of the existence of a policy π_∞ that could be viewed heuristically as the “optimal policy at infinity,” and the temporal homogeneity of this policy then gave us access to the machinery of Markov additive processes. There are many MDPs that offer similar prospects.

It took some specialized effort to relate the finite horizon policy π_n^* to the limiting policy, but the pattern used here offers some general guidance. In almost any MDP, the Bellman equation gives one good prospects for computing the value function, but here we benefitted most substantially from our understanding of the geometry of the threshold functions that determine the optimal policy. Our development of this understanding would have been stymied without the guidance provided by Figure 1. If one views our analysis as a case study, then one message is that given almost any MDP one would be wise to begin with the best numerical work that the problem allows.

Finally, the Bellman equation guarantees a natural rôle for induction in the analysis of many MDPs, but a more nuanced observation that emerges here is that it can be especially profitable to be attentive to various manifestations of supermodularity (or submodularity). Without the special supermodularity properties represented by (4) and (17) most of our inductions could not have moved forward. One can anticipate some aspect of this experience will be present in the analysis of a wide range of MDPs.

REFERENCES

- Arlotto, A., Chen, R. W., Shepp, L. A. and Steele, J. M. (2011), ‘Online selection of alternating subsequences from a random sample’, *J. Appl. Probab.* **48**(4), 1114–1132.
- Baryshnikov, Y. M. and Gnedin, A. V. (2000), ‘Sequential selection of an increasing sequence from a multidimensional random sample’, *Ann. Appl. Probab.* **10**(1), 258–267.
- Bateni, M., Hajiaghayi, M. and Zadimoghaddam, M. (2010), Submodular secretary problem and extensions, in ‘Approximation, randomization, and combinatorial optimization’, Vol. 6302 of *Lecture Notes in Comput. Sci.*, Springer, Berlin, pp. 39–52.
- Bertsekas, D. P. and Shreve, S. E. (1978), *Stochastic optimal control: the discrete time case*, Vol. 139 of *Mathematics in Science and Engineering*, Academic Press Inc. [Harcourt Brace Jovanovich Publishers], New York, NY.
- Bruss, F. T. and Delbaen, F. (2001), ‘Optimal rules for the sequential selection of monotone subsequences of maximum expected length’, *Stochastic Process. Appl.* **96**(2), 313–342.
- Bruss, F. T. and Delbaen, F. (2004), ‘A central limit theorem for the optimal selection process for monotone subsequences of maximum expected length’, *Stochastic Process. Appl.* **114**(2), 287–311.
- Buchbinder, N., Jain, K. and Singh, M. (2010), Secretary problems via linear programming, in ‘Integer programming and combinatorial optimization’, Vol. 6080 of *Lecture Notes in Comput. Sci.*, Springer, Berlin, pp. 163–176.
- Cayley, A. (1875), ‘Mathematical questions and their solutions’, *Educational Times* **22**, 18–19. See *The Collected Mathematical Papers of Arthur Cayley*, **10**, 587–588, (1986). Cambridge University Press, Cambridge.
- Dynkin, E. B. (1963), ‘Optimal choice of the stopping moment of a Markov process’, *Dokl. Akad. Nauk SSSR* **150**, 238–240.
- Gnedin, A. V. (1999), ‘Sequential selection of an increasing subsequence from a sample of random size’, *J. Appl. Probab.* **36**(4), 1074–1085.
- Gnedin, A. V. (2000a), ‘A note on sequential selection from permutations’, *Combin. Probab. Comput.* **9**(1), 13–17.

- Gnedin, A. V. (2000*b*), Sequential selection of an increasing subsequence from a random sample with geometrically distributed sample-size, in 'Game theory, optimal stopping, probability and statistics', Vol. 35 of *IMS Lecture Notes Monogr. Ser.*, Inst. Math. Statist., Beachwood, OH, pp. 101–109.
- Jones, G. L. (2004), 'On the Markov chain central limit theorem', *Probab. Surv.* **1**, 299–320.
- Krieger, A. M. and Samuel-Cahn, E. (2009), 'The secretary problem of minimizing the expected rank: a simple suboptimal approach with generalizations', *Adv. in Appl. Probab.* **41**(4), 1041–1058.
- Lindley, D. V. (1961), 'Dynamic programming and decision theory', *Appl. Statist.* **10**, 39–51.
- Lindvall, T. (2002), *Lectures on the coupling method*, Dover Publications Inc., Mineola, NY. Corrected reprint of the 1992 original.
- Meyn, S. and Tweedie, R. L. (2009), *Markov chains and stochastic stability*, second edn, Cambridge University Press, Cambridge. With a prologue by Peter W. Glynn.
- Samuels, S. M. (1991), Secretary problems, in 'Handbook of sequential analysis', Vol. 118 of *Statist. Textbooks Monogr.*, Dekker, New York, pp. 381–405.
- Samuels, S. M. and Steele, J. M. (1981), 'Optimal sequential selection of a monotone sequence from a random sample', *Ann. Probab.* **9**(6), 937–947.